Informatics in Primary Care 2012;20:95–102

Refereed paper

Can data extraction from general practitioners' electronic records be used to predict clinical outcomes for patients with type 2 diabetes?

Michael Staff MBBS MMedSc FAFPHM
Public Health Physician, Northern Sydney and Central Coast Local Health Districts, NSW Health and Clinical Senior Lecturer, School of Public Health, University of Sydney, Australia

ABSTRACT

Background The review of clinical data extraction from electronic records is increasingly being used as a tool to assist general practitioners (GPs) manage their patients in Australia. Type 2 diabetes (T2DM) is a chronic condition cared for primarily in the general practice setting that lends itself to the application of tools in this area.

Objective To assess the feasibility of extracting data from a general practice medical record software package to predict clinically significant outcomes for patients with T2DM.

Methods A pilot study was conducted involving two large practices where routinely collected clinical data were extracted and inputted into the United Kingdom Prospective Diabetes Study Outcomes Model to predict life expectancy. An initial assessment of the completeness of data available was performed and then for those patients aged between 45 and 64 years with adequate data life expectancies estimated.

Results A total of 1019 patients were identified as current patients with T2DM. There were sufficient data available on 40% of patients from one practice and 49% from the other to provide inputs into the UKPDS Outcomes Model. Predicted life expectancy was similar across the practices with women having longer life expectancies than men. Improved compliance with current management guidelines for glycaemic, lipid and blood pressure control was demonstrated to increase life expectancy between 1.0 and 2.4 years dependent on gender and age group.

Conclusion This pilot demonstrated that clinical data extraction from electronic records is feasible although there are several limitations chiefly caused by the incompleteness of data for patients with T2DM.

Keywords: clinical informatics, diabetes, life expectancy, primary care

What this paper adds

- Confirms the feasibility of accessing routinely collected electronic data on diabetes mellitus management in an Australian general practice setting.
- Demonstrates that estimates of health outcomes for people with type 2 diabetes mellitus can be generated using these data.
- Highlights the need for further research to fully realise the potential of applying computer models in a general practice setting to evaluate chronic disease management strategies.

Introduction

Type 2 diabetes (T2DM) has reached epidemic proportions in developed countries such as Australia, has become one of the most common non-communicable diseases in the world and results in substantial morbidity and mortality.[1] The disease is primarily managed in general practice and is often the focus of initiatives aimed at improving the care of patients with chronic medical conditions.[2,3]

With the increasing computerisation of general practice in Australia the use of clinical data extraction as a tool to assist practitioners manage their patients and participate in quality improvement activities has increased.[4] The Australian experience has largely focused on practice level compliance with clinical care parameters such as the proportion of patients with diabetes having a recent glycosylated haemoglobin (HbA1c) level below the target level.

General practice records can provide a very rich source of data that can be used to research chronic conditions such as T2DM provided ethical and privacy issues are adequately addressed.[5] With the advent and refinement of diabetes modelling software[6] this data potentially could be used to predict clinical outcomes for patients and be employed as a measure of the effectiveness of various chronic care interventions.

The aims of this pilot study are to assess the feasibility of extracting data suitable for use in modelling clinical outcomes from one of the more popular medical record software packages being used in Australian general practice, and to test whether these data can be successfully inputted into the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model[7] to predict clinically significant outcomes.

Methodology

Recruitment of general practitioners

Two general practices were identified among those who participated in a recently established general practitioner influenza surveillance programme that utilises electronic medical record data extractions.[8] Practices had to be fully computerised for clinical notes, use the Best Practice[9] medical record software (a common platform used in Australian general practice) and routinely receive patient laboratory results via Health Level seven (HL7) messaging to be eligible for inclusion in the pilot. Information technology terms are given in Box 1.

Best Practice software

This software provides a comprehensive electronic medical record capacity, uses a Microsoft SQL server database and is typically installed on a local network in the Australian general practice setting. Data on presenting complaints, clinical measurements, past history, social history and medication history are readily recorded both through the use of drop-down menus with automated coding and by free text recording. The software has a capacity to receive Health Language

Box 1 Information Technology Terms

SQL (Structured Query Language)
A standardised query language for requesting information from a database. It is an American National Standards Institute (ANSI) standard.

ODBC (Open Database Connectivity)
A standard database access method that allows access to data regardless of the database management system (DBMS) handling the data.

Microsoft SQL Server
A relational database server that was developed by Microsoft. It stores and retrieves data for other software applications to use.

HL7 (Health Level seven)
A standard for exchanging electronic information between medical applications. It defines a format for the transmission of electronic health-related information such as patient records, laboratory records and billing information.

CSV (Comma Separated Value)
A file format that stores numbers and text in a plain text form delineated by commas.

seven (HL7) messages from laboratories for automated inclusion of results into individual patient records.

Data extraction and manipulation

Data were extracted from participating general practices onsite using ODBCView, a free SQL query tool used to view and export data from any OBDC database.[10] This process generated CSV files that were manipulated by the statistical software program STATA[11] to produce the required input text file for the standalone version of the UKPDS Outcome Model software.[12]

Diabetes prediction modelling

The UKPDS Outcomes Model was used to predict major diabetes-related complications for type 2 diabetics (T2Ds) who attend the pilot practices.[7,12] T2Ds were identified through interrogation of both Best Practice internal codes and free text past history fields. There are multiple diagnosis field 'drop-down' options within Best Practice that are consistent with a diagnosis of T2D and these are internally assigned to either the internal code 374 or 378. Free text searching looked for the following terms; NIDDM, non-insulin diabetes mellitus, diabetes, type 2 and type II and excluded IDDM, insulin-dependent diabetes mellitus and type I. Patients who were recorded as having diabetes, but not categorised as insulin dependent or not, were excluded if they had a record of having been prescribed an exogenous insulin preparation. Predicted life expectancy is presented as the major outcome of interest. The minimum input dataset required to run the model and how these parameters were represented are described in Table 1.

Table 1 Input parameters for UKPDS Outcomes Model

Parameter	How calculated	Comment
Gender	From patient details	
Age at diagnosis	Calculated from date of birth	
Ethnicity	Defaulted to 'white'	Very poorly coded, default used
Duration of diabetes (years)	From 'past history' record	
Current clinical		
Glycosylated haemoglobin (%)	Average of calendar year measurements	'Current date' – '09 or '10
Smoking status (never, past, current)	From social history	
Cholesterol	Average of calendar year measurements	± 1 year from 'current date'
HDL	Average of calendar year measurements	± 1 year from 'current date'
Systolic blood pressure (mmHg)	Average of calendar year readings	± 1 year from 'current date'
Height (m)	From observation records	Any recording within 5 years of 'current date'
Weight (kg)	Average of calendar year measurements	± 1 year from 'current date'
Clinical at diagnosis		
Atrial fibrillation present	From 'past history' record	
Peripheral vascular disease present	From 'past history' record	
Cholesterol	Current reading if missing	
HDL	Current reading if missing	
Systolic blood pressure	Current reading if missing	
Glycosylated haemoglobin	7, 9 or 11% if missing	Sensitivity analysis

The study considered data recorded between 2001 and 2010 for patients aged 45–64 years and considered as 'current patients' as at 2011. For a patient to be included in the model there needed to be a record of the minimum dataset variables collected at least once in the calendar years 2009 or 2010.

Life expectancy was calculated in two ways; first, including the parameters listed in Table 1 only and second, including parameters in Table 1 plus projected yearly data from 2011 onwards. The projected data were set as either the upper level of target parameters or the current level (if that was lower) and smoking status was either maintained as non or past smoking or, for current smokers, set to past smoking. This was done in order to demonstrate the potential benefits that could be achieved by meeting recommended management guidelines. The target parameters used were; HbA1c 48 mmol/mol (6.5%), cholesterol 4.0 mmol/L, high-density lipoprotein (HDL) 1.0 and systolic blood pressure 140 mmHg.

UKPDS Outcome Modelling parameters

The model needs to deal with first-degree (Monte-Carlo variability) and second-degree (statistical variability) uncertainty.[13] First-degree uncertainty arises in the model because an individual's chances of developing one diabetes-related event may influence the risk of experiencing another in the future. The model allows for this by running multiple simulations that vary the order in which event equations are calculated.[7] Statistical variability is dealt with by applying a bootstrap procedure.[14] The UKPDS Outcomes Model user needs to specify both the number of Monte-Carlo trials and bootstraps used; in this study, 1000 trials and 1000 bootstraps were employed.[15]

Sensitivity analysis

The difficulty in obtaining clinical parameters at the time of the diagnosis of T2DM required the estimation of these parameters for many individuals. Where levels for cholesterol, HDL and systolic blood pressure at diagnosis were not recorded, they were estimated as the level recorded in 2009/10. A life expectancy sensitivity analysis was done to explore the impact of unknown HbA1c levels at diagnosis by substituting 7, 9 and 11% (53, 75 and 97 mmol/mol) as this level for patients without recorded levels. Where the HbA1c level at diagnosis was recorded this was used and no substitution performed.

Ethics

The study was approved by the Northern Sydney Health Service Ethics Committee and appropriate data custodian protocols implemented. Only non-identifiable data were extracted from medical records. The UKPDS Outcomes Model was used under academic licence from the University of Oxford, UK.

Results

Practice details

There were two practices included in this pilot study. Practice 1 had 17 full- or part-time general practitioners (GPs) and performed approximately 1500 consultations per week. Practice 2 had 5 full- or part-time GPs and performed 400 consultations per week. A total of 1019 current T2Ds were identified from the two practices. The median duration of diabetes differed between the practices with medians of 3 and 7 years for practice 1 and practice 2, respectively.

Completeness of input parameters

Table 2 describes the proportion of T2D records where sufficient data on clinical input parameters for the UKPDS Outcomes Model were able to be extracted. Forty percent of current T2Ds had the required inputs for modelling from practice 1 with a higher percentage of 49% available from practice 2 records.

Predicted life expectancy by UKPDS Outcomes Model

Table 3 presents the modelled life expectancies of T2Ds considered as a group for each general practice by 10-year age groups. Predicted life expectancy was similar between practices with females having longer life expectancies than males in general. Table 3 also demonstrates that the sensitivity analysis looking at the effect of varying the HbA1c at diagnosis had little effect on life expectancies. Forty-nine of the 176 (28%) modelled individuals had HbA1c levels at diagnosis recorded.

The effect of incorporating future annual clinical parameters equal to guideline levels is displayed in Table 4. It can be seen that between 1.0 and 2.4 years of additional years of life can be expected when comparing these levels with those estimated using default predicted clinical parameter levels included in the model.

Table 2 Type 2 diabetic patients with UKPDS Outcomes Model input parameters available from electronic their electronic medical record

Parameter	Practice 1 ($N = 881$)		Practice 2 (N = 138)	
	no.	%	no.	%
Glycosylated haemoglobin (%)	585	66.4	81	58.7
Cholesterol (mmol/L)	580	65.8	80	58.0
HDL cholesterol (mmol/L)	563	63.9	80	58.0
Systolic blood pressure (mmHg)	535	60.7	80	58.0
Smoking status (never, past, current)	715	81.2	131	94.9
Height (m)	574	65.2	126	91.3
Weight (kg)	472	53.6	71	51.4
All required parameters	349	39.6	67	48.6

Table 3 Modelled life expectancy, sensitivity analysis for glycosylated haemoglobin levels (HbA1c) at diagnosis of type 2 diabetes

	Practice 1				Practice 2			
Age group (years)*	No.	Diagnosis HbA1c (%)	Life expectancy (years)	95% CI	No.	Diagnosis HbA1c (%)	Life expectancy (years)	95% CI
Males								
45–54	29	7	22.5	(19.9, 25.0)	4	7	20.9	(18.4,23.5)
		9	22.2	(19.7, 24.7)		9	20.6	(18.1, 23.1)
		11	22.0	(19.6, 24.4)		11	20.8	(18.4, 23.2)
55–64	62	7	16.4	(14.9, 18.0)	12	7	15.8	(14.0, 17.5)
		9	16.3	(14.7, 17.8)		9	15.7	(14.0, 17.4)
		11	16.2	(14.7, 17.7)		11	15.5	(13.8, 17.1)
Females								
45–54	19	7	23.4	(20.5, 26.3)	5	7	23.6	(20.5, 26.7)
		9	23.3	(20.4, 26.1)		9	23.4	(20.3, 26.5)
		11	23.1	(20.3, 25.9)		11	23.2	(20.1, 26.2)
55–64	35	7	17.9	(16.0, 19.8)	10	7	19.1	(17.1, 21.1)
		9	17.8	(15.9, 19.6)		9	18.8	(16.9, 20.8)
		11	17.6	(15.8, 19.4)		11	18.7	(16.8, 20.6)

* Based on age of patient in 2009 or 2010.

Discussion

This pilot study found that it was feasible to identify the electronic medical records of patients with T2D produced by a software package commonly used within Australia general practice and extract data to predict life expectancy and complications for this group of patients. Although somewhat disappointing, 40–50% of current T2D patients' records provided sufficient model inputs to allow meaningful predictions to be made. The estimates of life expectancies

Table 4 Modelled life expectancy using predicted annual clinical parameters based on current management guidelines for type 2 diabetes

	Practice 1				Practice 2			
Age group (years)*	No.	Life expectancy (years)	95% CI	Difference (years)†	No.	Life expectancy (years)	95% CI	Difference (years)†
Male								
45–54	29	24.4	(21.4, 27.5)	2.4	4	22.8	(19.7, 15.9)	2.2
55–64	62	18.0	(16.1, 20.0)	1.8	12	17.4	(15.2, 19.6)	1.9
Female								
45–54	19	24.2	(21.0, 27.4)	1.1	5	24.2	(20.8, 27.6)	1.0
55–64	35	18.9	(16.7, 21.1)	1.7	10	19.9	(17.6, 22.2)	1.2

* Based on age of patient in 2009 or 2010. # Life expectancy in Table 4 – life expectancy in Table 3.

made are similar to those reported by UKPDS Group when they applied the model to the UKPDS population.[7]

The findings of this pilot study are important because they demonstrate that routinely collected clinical data can be used to directly estimate the effectiveness of diabetes management strategies in terms of clinical outcomes. Doing so removes the need to rely on the assumption that meeting prescribed guidelines and targets will result in significant clinical benefit.

A recent literature review identified that information technology has been successfully used to provide clinicians with data on both process and clinical measures of diabetes care in the primary care setting. These data have, in turn, been used to enhance healthcare delivery.[16] A variety of strategies have been used including the use of diabetic patient registers, tracking HbA1c tests, point of care decision support tools and clinician reminders.[17–19] These strategies have mainly been evaluated in a quality improvement framework and in some cases required the augmentation of existing electronic medical record systems.[17]

While it is acknowledged that data extraction tools can be very useful for identifying, quantifying and monitoring clinical practice issues, considerable technical barriers have been identified that potentially limit their usefulness.[2,20] These include some shortcomings of data extraction tools,[20] variation in recording of data,[21] comprehensive data entry[20] and acceptance of practice staff.[22] However, even given these limitations others have still reported that roughly half of patients with T2DM have recent data available for HbA1c, cholesterol and systolic blood pressure in their electronic record.[20]

Despite there being a variety of software and vendors across the healthcare industry in Australia the widespread use of the HL7 standard has enabled near real-time electronic delivery of pathology results.[23] It is likely that this has already resulted in a larger amount of clinical data relevant to diabetes management being available to GPs. Developments such as this along with continued interest and support in quality improvement initiatives based upon the use of electronic data recording[2,4] should help overcome the barriers identified above.

The UKPDS Outcomes Model requires considerable individual-level clinical data to be inputted in order to make predictions. A major issue highlighted by the current study is that data on clinical parameters such as HbA1c at the time of diabetes diagnosis are unlikely to be routinely available. The developers of the software have anticipated this and suggested two options to overcome this problem.[24] First, to conduct a sensitivity analysis (as done in this study) for some variables or, second, to predict risk factors directly and input these as a series of future annual data. The sensitivity analysis done in this study demonstrated that the paucity of data on HbA1c at diagnosis had little impact on estimated life expectancies.

The inclusion of projected risk factor data based upon management guideline targets demonstrated that improved management of patients could achieve up to an additional 2.4 years of life expectancy. This benefit, important in itself, may be even more so if the patients included in the modelling were those with better disease management and control.

There are several limitations to the current study. First, it is a pilot and as such requires the participation

of a greater number of practices to confirm its findings. Second, outcomes were predicted for only those individuals with a complete baseline dataset and it is reasonable to suspect that these individuals were more likely to present for routine care on a regular basis. Consequently, they may well have better control of their disease than the other diabetic patients from the practice and the outcomes reported at a practice level such as life expectancy may well be an overestimate due to selection bias.

The UKPDS Outcomes Model was developed using the data from the United Kingdom Prospective Diabetes Study[25] and has yet to be validated in an Australian setting. There have been attempts to test cardiovascular risk equations derived from this study in Australian patients with type 2 diabetes using data collected as part of the Fremantle Diabetes Study.[26] The authors attempted to validate the UKPDS risk engine among 791 T2Ds recruited from a single urban centre and concluded that its underlying equations are not suitable for predicting risk in Australians. It should, however, be emphasised that the UKPDS Outcomes Model differs from the UKPDS risk engine in that it has considerably more input data, includes information on previous events and can take account of updated risk factor data over time.

Another limitation of the study is related to the ability of the UKPDS Outcomes Model to predict life expectancy at an individual level. Although the model reports predictions at an individual level it should be emphasised that the uncertainty of an estimated life expectancy for any individual patient is likely to be substantial.[7] As such more reliance should be placed on practice-level predictions which are much more robust.

This pilot study raises the need for further research. It is important that its findings are replicated on a larger scale and when using other medical software packages. Further research into the barriers faced by GPs when challenged to improve the quality of electronic medical and how these could be overcome should be encouraged.

Data extraction from electronic medical records routinely used in the Australian general practice setting can be used to predict clinical outcomes among T2D patients at a practice level. The outcomes presented in this pilot may be limited by selection bias and the validity of the underlying model upon which they are based is yet to be established in an Australian setting. Nevertheless, this pilot study highlights how routinely collected data could be used to estimate the clinical impact of diabetes management strategies in the general practice setting.

ACKNOWLEDGEMENTS

I would like to thank the general practitioners who allowed access to their medical record systems to provide data for this study. I would also like to thank Professor Lyn March for her kind comments on a draft of this article.

REFERENCES

1 Barr ELM, Magliano DJ, Zimmet PZ *et al. AusDiab 2005: The Australian Diabetes, Obesity and Lifestyle Study. Tracking the accelerating epidemic: its causes and outcomes.* Melbourne: International Diabetes Institute, 2006. www.bakeridi.edu.au/Assets/Files/AUSDIAB_REPORT_2005.pdf
2 Australian Primary Care Collaborative. Improvement Foundation. www.apcc.org.au/about_the_APCC/program_results
3 *Diabetes Management in General Practice* (15e) 2009/10. Diabetes Australia 2009. Publication NP 1055 ISBN 978 1875 690 190. www.racgp.org.au/diabetes/systemforcare
4 Ford D and Knight A. The Australian Primary Care Collaborative: an Australian general practice success story. *Medical Journal of Australia* 2010;193(2):90–1.
5 Mathers N, Perrin N and Watt G. Using patient records from general practice for research. *Informatics in Primary Care* 2009;17:137–9.
6 The Mount Hood 4 Modeling Group. Computer modeling of diabetes and its complications. *Diabetes Care* 2007;30(6):1638–46.
7 Clarke P, Gray A, Briggs A *et al.* A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia* 2004;47:1747–59.
8 Liljeqvist G, Staff M, Puech M, Blom H and Torvaldsen S. Automated data extraction from general practice records in an Australian setting: trends in influenza-like illness in sentinel general practices and emergency departments. *BMC Public Health* 2011;11:435. http://www.biomedcentral.com/1431–2458/11/435
9 Best Practice Software. www.bpsoftware.com.au (accessed 01/12).
10 ODBC View. Slik Software Ltd. www.sliksoftware.co.nz/products/odbcview (accessed 01/12).
11 StataCorp. 2008. *Stata Statistical Software: Release 10.* College Station, TX: StataCorp LP.
12 UKPDS Outcomes Model. version 1.3 © Isis Innovation Ltd 2010 www.dtu.ox.ac.uk/outcomesmodel (accessed 01/12).
13 American Diabetes Association Consensus Panel. Guidelines for computer modeling of diabetes and its complications. *Diabetes Care* 2004;27(9):2262–5.
14 Campbell MK and Torgerson DJ. Bootstrapping: estimating confidence intervals for cost-effectiveness ratios. *QJM* 1999;92:177–82.
15 UKPDS Outcomes Model User manual. version 1.3, January 2011. ISIS Innovation Ltd. www.dtu.ox.ac.uk/

outcomesmodel/UKPDSOutcomesManual.pdf (accessed 01/12).
16 Adaji A, Schnattner P and Jones K. The use of information technology to enhance diabetes management in primary care: a literature review. *Informatics in Primary Care* 2008;16(3):229–37.
17 Hunt JS, Siemienczuk J, Gillanders W *et al.* The impact of a physician-directed health information technology system on diabetes outcomes in primary care: a pre and post implementation study. *Informatics in Primary Care* 2009;17(3):165–74.
18 Chaudry R, Tutlegde-Scheitel SM, Thomas MR *et al.* Clinical informatics to improve quality of care: a population based system for patients with diabetes mellitus. *Informatics in Primary Care* 2009;17(2):95–102.
19 Howard JA, Sommers R, Gould ON and Mancuso M. Effectiveness of an HbA1c tracking tool on primary care management of diabetes mellitus: glycaemic control, clinical practice and usability. *Informatics in Primary Care* 2009;17(1):41–6.
20 Schattner P, Saunders M, Stranger L, Speak M and Russo K. Clinical data extraction and feedback in general practice: a case study from Australian primary care. *Informatics in Primary Care* 2010;18:205–12.
21 Rollason W, Khunti K and de Lusignan S. Variation in the recording of diabetes diagnostic data in primary care computer systems: implications for quality of care. *Informatics in Primary Care* 2009;17(2):113–19.
22 Civil M, Bulsara C, Karner S *et al.* Attitudes and practices of recording diabetic patient information within an Australian general practice setting: an exploratory study. *Informatics in Primary Care* 2009;17(1):35–9.
23 ehealth Australia. http://ehealthaustralia.org/article/healthcare-health-informatics-and-interoperability-in-australia (accessed 01/12).
24 www.dtu.ox.ac.uk/outcomesmodel/FAQ.php (accessed 01/12).
25 UKPDS Group. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *BMJ* 2000;321:405–12.
26 Davis W, Colagiuri S and Davis M. Comparison of the Framingham and United Kingdom Prospective Diabetes Study cardiovascular risk equations in Australian patients with type 2 diabetes from the Fremantle Diabetes Study. *Medical Journal of Australia* 2009;190(4):180–4.

ADDRESS FOR CORRESPONDENCE

Michael Staff
Public Health Unit
Hornsby Hospital
Palmerston Rd
Hornsby
NSW 2077
Australia
Email: mstaff@nsccahs.health.nsw.gov.au

Accepted March 2012